



Huma-Num

la TGIR des humanités numériques



Journées

PRÉSERVER SES DONNÉES NUMÉRIQUES

6-7 FÉVRIER 2020 - CAMPUS CONDORCET



Journées Préserver ses données numériques

Huma-Num : la préparation des données

Helene Jouguet
TGIR huma-Num UMS 3598
Licence CC-BY



Aix-Marseille
université

CAMPUS
CONDORCET
Paris-Aubervilliers

Développer la qualité des données



Des données complètes, à jour, disponibles et utilisables

Une réflexion et des choix pour la gestion des données tout au long du projet

Un ensemble d'action qui permettent :

de prémunir les données de la dégradation, de la perte physique, de la perte de sens etc.

de rendre visible les données pour leur re-exploitation

Développer la qualité des données



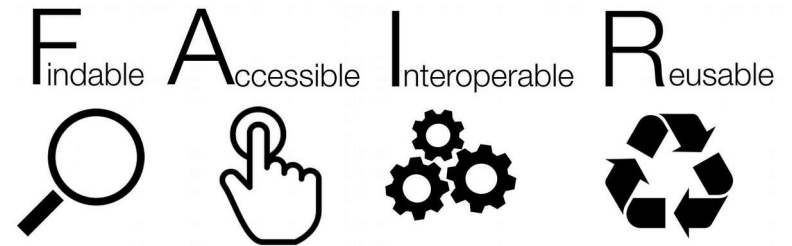
Auditer les porteurs de projet pour les conseiller et les orienter

Suivre la vie de leur projet et de leurs données dans nos services

Développer la formation et les supports de formations aux services d'Huma-Num

Veiller à la mise en œuvre des principes FAIR sur les données hébergées

Mise en conformité avec les principes F. A. I. R



Leviers :

Des services numériques offrant les fonctionnalités requises

Un travail documentaire sur les données de la part des producteurs

Des processus organisationnels sur le long terme

Données FAIR = Données préservées

Ouverture des données

OUVRIR LA SCIENCE !

Loi république numérique, plan national pour la science ouverte,
H2020, ANR : Exigence d'ouverture des données forte

Exposition de données ouvertes = forme de préservation

On ne peut pas se permettre d'ouvrir ses données sans qu'elles soient
bien documentées

<https://www.ouvrirlascience.fr/plan-de-gestion-de-donnees-recommandations-a-lanr/>

Briques du circuit des données



Tout au long du projet : différents acteurs / interlocuteurs

Huma-Num s'inscrit dans la chaîne de vie des données : services et outils pour procéder à la création, collecte, traitement, documentation et exposition des données de votre projet

Services d'Huma-Num

Planification
Organisation

Collecte
Création

Analyse
Traitement

Archivage
Conservation

Partage
Publication

Découverte
Réutilisation

Outils et
programmes

 ShareDocs

 nakala

 isidore

 GitLab

 nakalona

 DIFFUSER

Data Management Plan

 Huma-Num

Services d'Huma-Num

Planification
Organisation

Collecte
Création

Analyse
Traitement

Archivage
Conservation

Partage
Publication

Découverte
Réutilisation

Outils et
programmes

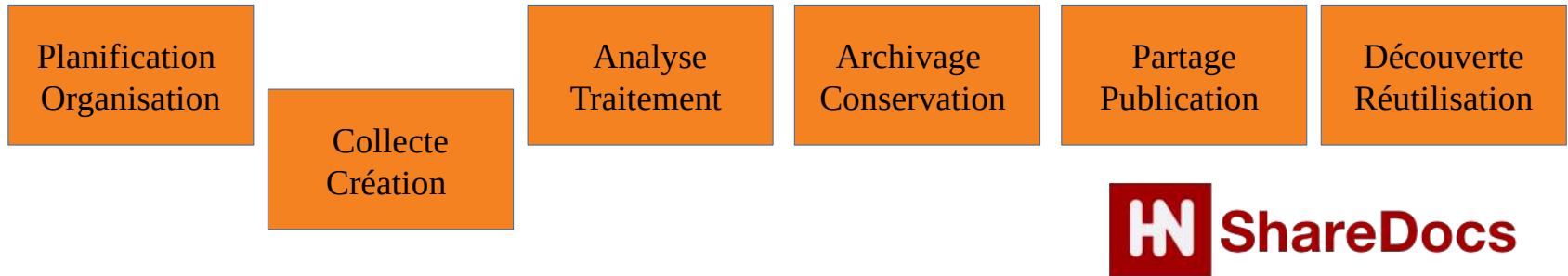
Les outils sont indispensables

La façon de les utiliser va garantir la préservation des données

Accompagner l'utilisation des services qu'Huma-Num met à disposition

Chaque service nécessite des méthodes et actions qui permettront la mise en œuvre de cette qualité

Data Management Plan



Le service de stockage sécurisé

Fonction

garantit la conservation et l'intégrité de fichiers informatiques.
offre des fonctionnalités pour le travail collaboratif (accès en ligne, gestion des droits)

Préservation : **Assurer de comprendre et d'identifier les données**

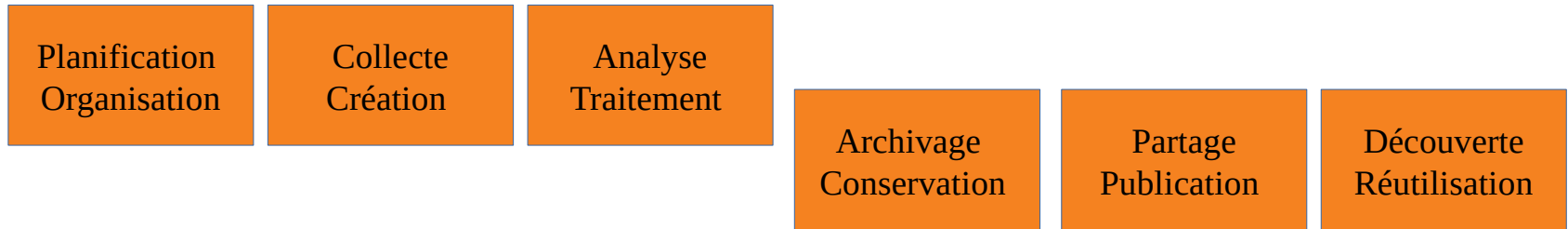
Organiser les données : plan de classement

Identifier les données : définition d'un plan nommage des fichiers

Réseau Qualité en Recherche (CNRS)

Traçabilité des activités de recherche et gestion des connaissances - Guide pratique de mise en place :

<http://qualite-en-recherche.cnrs.fr/spip.php?article315>



Collecte, Archivage, Partage : la description des données

Rôle des métadonnées

Nécessité d'utiliser des métadonnées standardisées

<https://doranum.fr/metadonnees-standards-formats/fiche-synthetic/>

Nakala : Dublin Core

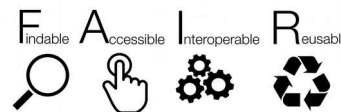


Fonction : Mettre à disposition des communautés SHS un outil complet garantissant un accès sécurisé, persistant et interopérable à leurs données.

Préservation des données : une description normalisée et « riche »

=> F2 Les métadonnées décrivant les données sont riches.

=> I1 Les données et les métadonnées utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances.



Description Dublin Core



Dublin Core™ Metadata Initiative

Permettre une interopérabilité minimale entre des systèmes d'origines différentes.

Nakala, 18 éléments facultatifs et répétables :

Description du contenu :

Title, Subject, Description, Source, Language, Relation, Coverage ;

Propriété intellectuelle :

Creator, Contributor, Publisher, Rights ;

Instanciation :

Date, Type, Format, Identifier.

Éléments du Dublin core qualifié :

Audience, provenance, rightsolder

Qualificatifs utilisables.

Description Dublin Core qualifié

rdf:type	<ul style="list-style-type: none"> foaf:Document 	rdf:type	<ul style="list-style-type: none"> foaf:Document
dcterms:title	<ul style="list-style-type: none"> Detailed dataset description.xlsx 	dcterms:title	<ul style="list-style-type: none"> Carte de Ambohimalaza P47II (fr)
dcterms:created	<ul style="list-style-type: none"> 19/06/2018 	dcterms:identifiant	<ul style="list-style-type: none"> FR_R_MADA_D09_11_00002
dcterms:creator	<ul style="list-style-type: none"> Nesrine Tarhouni 	dcterms:contributor	<ul style="list-style-type: none"> Alessandri Chanal Fournioux Fusy Noël Sayous Vincensini
Is foaf:primaryTopic of	<ul style="list-style-type: none"> <https://www.nakala.fr/resource/11280/4611> 	dcterms:coverage	<ul style="list-style-type: none"> Ambohimalaza (fr) Antananarivo (fr) Tananarive (fr)
dcterms:type	<ul style="list-style-type: none"> Excel file 	dcterms:created	<ul style="list-style-type: none"> 1950 (dcterms:W3CDTF)
		dcterms:creator	<ul style="list-style-type: none"> Service géographique de Madagascar
		dcterms:description	<ul style="list-style-type: none"> Dressé, dessiné et publié par le Service Géographique de Madagascar en 1933 Révisé
		dcterms:extent	<ul style="list-style-type: none"> 1:50.000
		dcterms:format	<ul style="list-style-type: none"> 62 x 43
		dcterms:language	<ul style="list-style-type: none"> fr (dcterms:RFC3066) français (fr) <http://lexvo.org/id/iso639-3/fra>
		dcterms:license	<ul style="list-style-type: none"> <https://creativecommons.org/licenses/by-nc-sa/4.0/>
		dcterms:medium	<ul style="list-style-type: none"> Méridien de Paris (fr) Papier (fr) Polychrome (fr) Projection conforme Laborde (fr)
		Is foaf:primaryTopic of	<ul style="list-style-type: none"> <https://www.nakala.fr/resource/11280/73b47d74>
		dcterms:spatial	<ul style="list-style-type: none"> east=47.6266; north=-18.9; name=Madagascar (dcterms:Point) east=47.6266; north=-18.9; name=Point central de la carte (dcterms:Point) northlimit=-18.8; southlimit=-19; westlimit=47.5532; eastlimit=47.7 (dcterms:Box)
		dcterms:subject	<ul style="list-style-type: none"> 1:50.000 bois (fr) habitat dispersé (fr) rivière (fr) rizière (fr)
		dcterms:type	<ul style="list-style-type: none"> Carte topographique (fr) <http://purl.org/dc/dcmitype/StillImage>

Dublin Core

<https://www.bnf.fr/fr/dublin-core>

Vocabulaires

Exemple : remplir le champs dc:subject : s'accorder dans l'équipe pour utiliser une norme.

Corpus linguistique : décrire le champ linguistique =
vocabulaire OLAC

Corpus archéologique : décrire les peuples = thésaurus
PACTOLS

2. Linguistic Subject

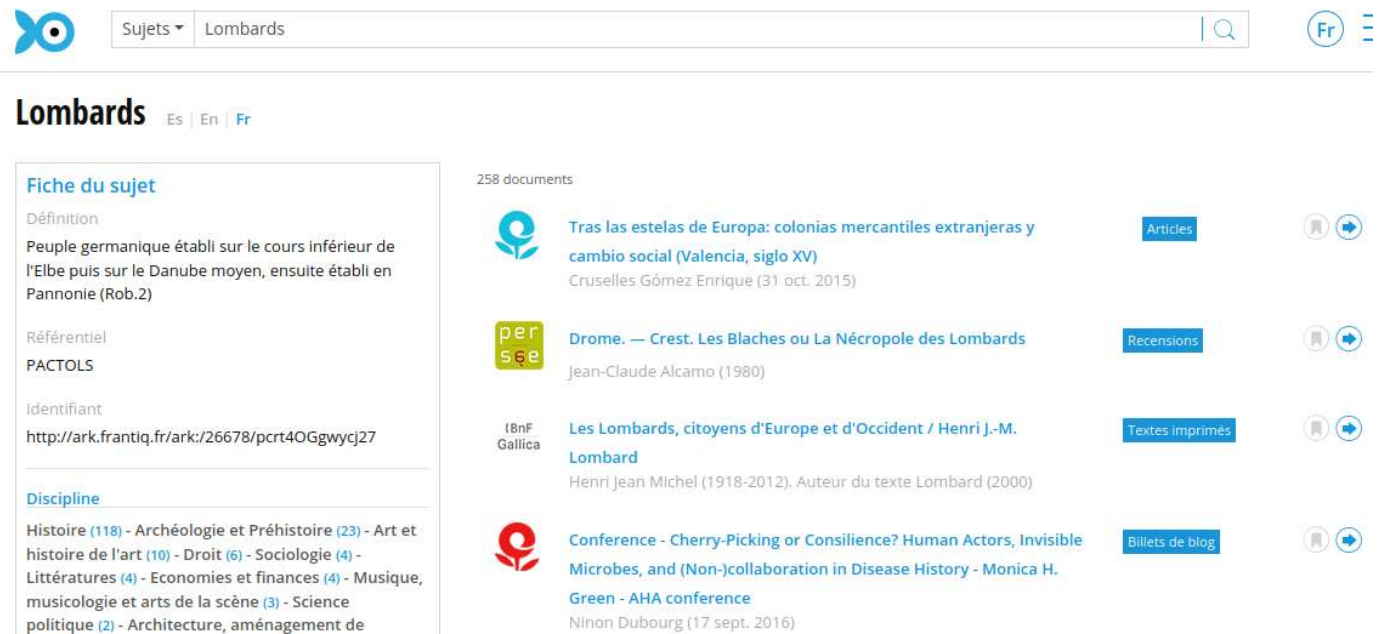
- anthropological_linguistics
- applied_linguistics
- cognitive_science
- computational_linguistics
- discourse_analysis
- forensic_linguistics
- general_linguistics
- historical_linguistics
- history_of_linguistics
- language_acquisition
- language_documentation
- lexicography
- linguistics_and_literature
- linguistic_theories
- mathematical_linguistics
- morphology
- neurolinguistics
- philosophy_of_language
- phonetics
- phonology
- pragmatics
- psycholinguistics
- semantics
- sociolinguistics
- syntax
- text_and_corpus_linguistics
- translating_and_interpreting
- typology
- writing_systems

The screenshot displays the PACTOLS interface. On the left is a hierarchical tree view with categories like 'Anthroponymes', 'Chronologie', 'Lieux', 'Oeuvres', and 'Peuples'. The 'Peuples' category is expanded, showing sub-categories such as 'Cultures préhistoriques', 'Africains', 'Amérindiens', 'Arabes', 'Asiatiques', 'Asiatiques', 'Australiens', 'Autres Peuples', 'Barbares', 'Celts', 'Gallo-romains', 'Gaulois', 'Germaniques', 'Grecs', 'Indo-européens', 'Indo-iraniens', 'Peuples légendaires', 'Peuples méditerranéens', 'Sémites', 'Vikings', and 'Peuples actuels'. The main content area shows the details for the term 'Germaniques', including its URI, number of notices (55), and various identifiers. Below this, there are buttons for 'Notes', 'Alignement', and 'Coordonnées Gps'. On the right side, there is a 'Facette' section with a list of languages and a 'Image' section with 'Images externes'.

<http://www.language-archives.org/documents.html#Recommendations>
<https://pactols.frantiq.fr/opentheso/>

Vocabulaires : exposition des données

Isidore : notices liées au terme Lombards (peuple) issu du thésaurus Pactols



The screenshot shows the Isidore search interface. At the top, there is a search bar with 'Sujets' and 'Lombards' selected. Below the search bar, the results for 'Lombards' are displayed. On the left, there is a 'Fiche du sujet' (Subject Card) for 'Lombards'. The card includes a definition: 'Peuple germanique établi sur le cours inférieur de l'Elbe puis sur le Danube moyen, ensuite établi en Pannonie (Rob.2)'. It also lists the 'Référentiel' as 'PACTOLS' and provides an 'Identifiant' (URI: http://ark.frantlq.fr/ark:/26678/pcrt4OGgwyj27). Below the card, a 'Discipline' section lists various fields of study with associated counts: Histoire (118), Archéologie et Préhistoire (23), Art et histoire de l'art (10), Droit (6), Sociologie (4), Littératures (4), Economies et finances (4), Musique, musicologie et arts de la scène (3), Science politique (2), and Architecture, aménagement de. On the right, a list of 258 documents is shown. The first three documents are: 1. 'Tras las estelas de Europa: colonias mercantiles extranjeras y cambio social (Valencia, siglo XV)' by Cruselles Gómez Enrique (31 oct. 2015), categorized as 'Articles'. 2. 'Drome. — Crest. Les Blaches ou La Nécropole des Lombards' by Jean-Claude Alcamo (1980), categorized as 'Recensions'. 3. 'Les Lombards, citoyens d'Europe et d'Occident / Henri J.-M. Lombard' by Henri Jean Michel (1918-2012), Auteur du texte Lombard (2000), categorized as 'Textes imprimés'. The fourth document is 'Conference - Cherry-Picking or Consilience? Human Actors, Invisible Microbes, and (Non-)collaboration in Disease History - Monica H. Green - AHA conference' by Ninon Dubourg (17 sept. 2016), categorized as 'Billets de blog'. Each document entry includes a small icon, the title, author, date, and category, along with icons for bookmarking and sharing.

Lombards Es | En | Fr

Fiche du sujet

Définition
Peuple germanique établi sur le cours inférieur de l'Elbe puis sur le Danube moyen, ensuite établi en Pannonie (Rob.2)

Référentiel
PACTOLS

Identifiant
http://ark.frantlq.fr/ark:/26678/pcrt4OGgwyj27

Discipline

Histoire (118) - Archéologie et Préhistoire (23) - Art et histoire de l'art (10) - Droit (6) - Sociologie (4) - Littératures (4) - Economies et finances (4) - Musique, musicologie et arts de la scène (3) - Science politique (2) - Architecture, aménagement de

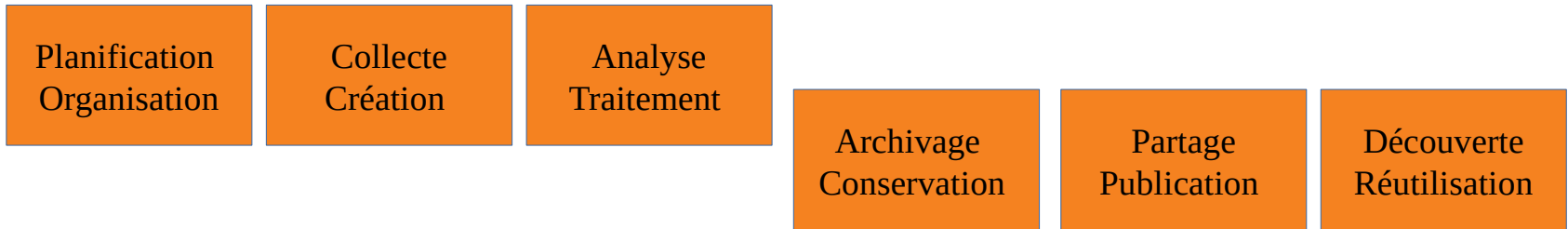
258 documents

Tras las estelas de Europa: colonias mercantiles extranjeras y cambio social (Valencia, siglo XV) Articles
Cruselles Gómez Enrique (31 oct. 2015)

Drome. — Crest. Les Blaches ou La Nécropole des Lombards Recensions
Jean-Claude Alcamo (1980)

Les Lombards, citoyens d'Europe et d'Occident / Henri J.-M. Lombard Textes imprimés
Henri Jean Michel (1918-2012), Auteur du texte Lombard (2000)

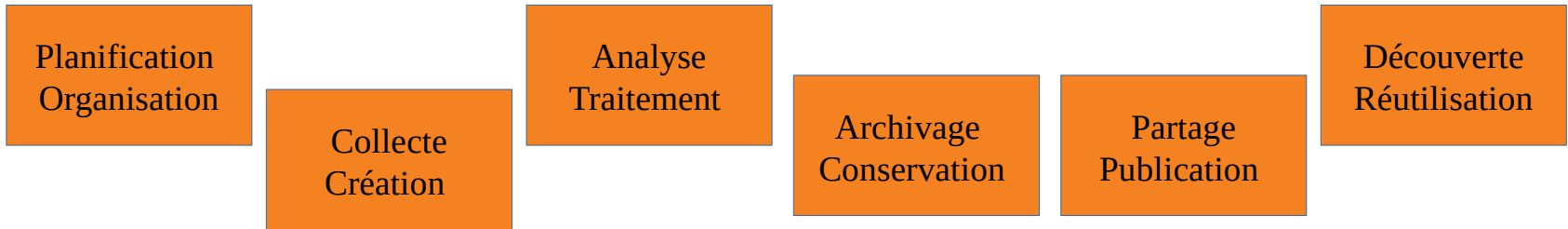
Conference - Cherry-Picking or Consilience? Human Actors, Invisible Microbes, and (Non-)collaboration in Disease History - Monica H. Green - AHA conference Billets de blog
Ninon Dubourg (17 sept. 2016)



Evolution de Nakala

Travaux en cours pour explorer la mise en œuvre d'un circuit automatisé entre Nakala et la plateforme d'archivage – PAC - du CINES

Intégration d'un dispositif de préparation des données spécifique pour répondre aux exigences de la plateforme d'archivage



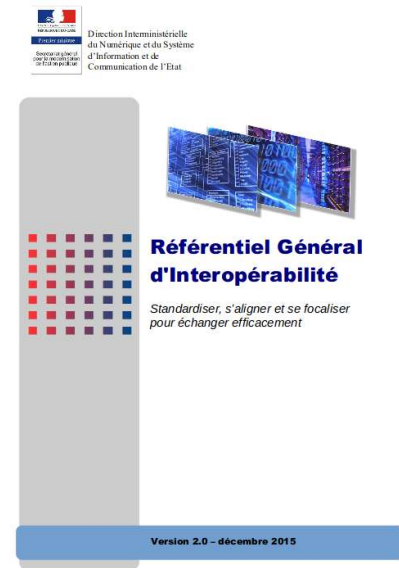
Collecte, Archivage, Partage : le choix des formats

Différents formats pour une même données ?

Acteurs spécialistes des formats

Cines : FACILE - Service de validation de formats

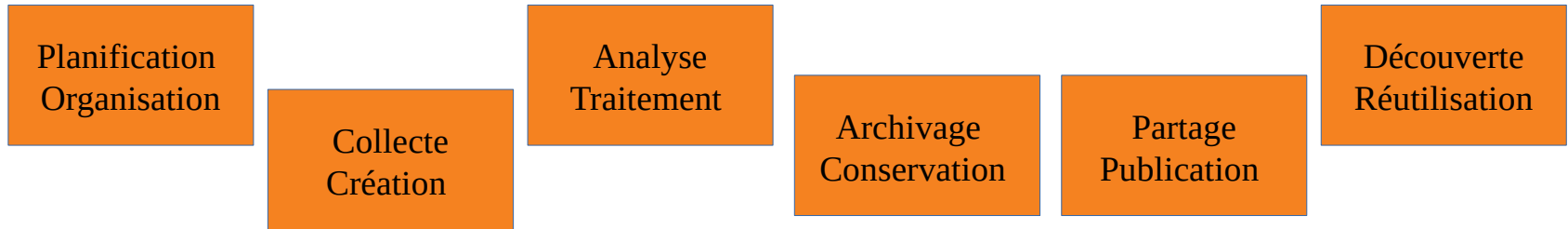
Communautés de recherche



http://references.modernisation.gouv.fr/sites/default/files/Referentiel_General_Interoperabilite_V2.pdf

<https://www.cines.fr/actualites/la-gazette-du-cines/>

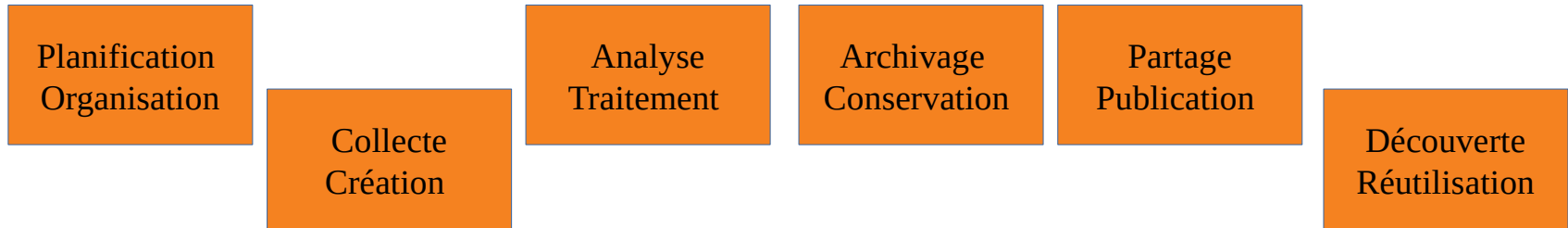
<https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/DANSpreferredformatsUK.pdf>



Collecte, Archivage, Partage : la sélection des données

Sélection des données selon les objectifs poursuivis ?

http://qualite-en-recherche.cnrs.fr/IMG/pdf/guide_tracabilite_activites_recherche_gestion_connaissances.pdf
https://repository.jisc.ac.uk/7262/1/JR0100_WHAT_RESEARCH_DATA_TO_KEEP_FEB2019_v5_WEB.pdf
https://www.archivistes.org/IMG/pdf/referentiel_recherche_intro_septembre2012_corrige_.pdf
<https://nerc.ukri.org/research/sites/data/policy/data-value-checklist/>



Collecter et préparer les données en respectant les principes éthiques et le RGPD

Préparer la réutilisation des données :

Notice DC : l'élément Rights permet de déclarer les droits applicables à la ressource décrite.

Licences de réutilisation : <https://www.data.gouv.fr/fr/licences>

<https://doranum.fr/aspects-juridiques-ethiques/les-licences-de-reutilisation-dans-le-cadre-de-lopen-data-2/>
InSHS Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte. GUIDE POUR LA RECHERCHE

Guide des bonnes pratiques éthiques et juridiques : <https://ethiquedroit.hypotheses.org/2970>



Huma-Num

la TGIR des humanités numériques



Journées

PRÉSERVER SES DONNÉES NUMÉRIQUES

6-7 FÉVRIER 2020 - CAMPUS CONDORCET



Journées Préserver ses données numériques

Huma-Num : la préparation des données

Helene Jouguet
TGIR huma-Num UMS 3598
Licence CC-BY



Aix-Marseille
université

CAMPUS
CONDORCET
Paris-Aubervilliers